



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

**A SAMPLING METHODOLOGY FOR INSURV
MATERIAL INSPECTIONS**

by

Ronald D. Fricker, Jr.

March 2012

Approved for public release; distribution is unlimited

Prepared for: President, Board of Inspection and Survey
2600 Tarawa Court, Suite 250
Virginia Beach, VA 2345

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 06-03-2012		2. REPORT TYPE Technical Report		3. DATES COVERED (From-To) 1 JUL 2011 – 30 SEP 2011
4. TITLE AND SUBTITLE A Sampling Methodology for INSURV Material Inspections			5a. CONTRACT NUMBER	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Ronald D. Fricker, Jr.			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Operations Research Department Graduate School of Operational and Information Science Naval Postgraduate School, Monterey, CA 93943			8. PERFORMING ORGANIZATION REPORT NUMBER N P S - O	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) President, Board of Inspection and Survey 2600 Tarawa Court, Suite 250 Virginia Beach, VA 2345			10. SPONSOR/MONITOR'S ACRONYM(S)	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT <p>This technical report summarizes research into sampling methods that the U.S. Navy Board of Inspection and Survey (INSURV) could employ during Material Inspections (MIs) of ships and submarines. The goal is to improve the Board's efficiency in conducting MIs without compromising either Board effectiveness or inspection rigor.</p> <p>The idea of sampling is that, rather than inspecting every item of a specific type—for example, portable CO2 bottles—onboard a ship or submarine, INSURV will only inspect a sample of those items. From the sample, INSURV would then characterize the ship's or submarine's entire complement of that item.</p> <p>This report outlines a sampling methodology that is statistically rigorous and therefore quantitatively defensible, and it is implementable. It is based on well-known sampling methods, such as those described in Cochran (1977) and Lohr (1999).</p> <p>The method described herein allows INSURV to specify the desired margin of error of the results on each item. It is expected that this decision will be based on the mission essentiality and/or safety criticality of each item, where items that are mission essential or safety critical will be given very small margins of error. Similarly, items that are not mission essential or safety critical will be given appropriately larger margins of error.</p>				
15. SUBJECT TERMS Sampling, Board of Inspection and Survey, Material Inspection, United States Navy				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 59
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified		
				19a. NAME OF RESPONSIBLE PERSON Ronald D. Fricker
				19b. TELEPHONE NUMBER (include area code) (831) 656-3048

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

Daniel T. Oliver
President

Leonard A. Ferrari
Executive Vice President and
Provost

The report entitled “*A Sampling Methodology for INSURV Material Inspections*” was prepared for and funded by the President, Board of Inspection and Survey, 2600 Tarawa Court, Suite 250, Virginia Beach, Virginia 2345.

Further distribution of all or part of this report is authorized.

This report was prepared by:

Ronald D. Fricker, Jr.
Associate Professor
Department of Operations Research

Reviewed by:

Robert F. Dell
Chairman
Department of Operations Research

Released by:

Douglas Fouts
Interim Vice President and
Dean of Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This technical report summarizes research into sampling methods that the U.S. Navy Board of Inspection and Survey (INSURV) could employ during Material Inspections (MIs) of ships and submarines. The goal is to improve the Board's efficiency in conducting MIs without compromising either Board effectiveness or inspection rigor.

The idea of sampling is that, rather than inspecting every item of a specific type—for example, portable CO2 bottles—onboard a ship or submarine, INSURV will only inspect a sample of those items. From the sample, INSURV would then characterize the ship's or submarine's entire complement of that item.

This report outlines a sampling methodology that is statistically rigorous and therefore quantitatively defensible, and it is implementable. It is based on well-known sampling methods, such as those described in Cochran (1977) and Lohr (1999).

The method described herein allows INSURV to specify the desired margin of error of the results on each item. It is expected that this decision will be based on the mission essentiality and/or safety criticality of each item, where items that are mission essential or safety critical will be given very small margins of error. Similarly, items that are not mission essential or safety critical will be given appropriately larger margins of error.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

This technical report, in the form of an annotated briefing, summarizes research into sampling methods that the U.S. Navy Board of Inspection and Survey (INSURV) could employ during Material Inspections (MIs) of ships and submarines. The goal is to improve the Board's efficiency in conducting MIs without compromising either Board effectiveness or inspection rigor.

The idea of sampling is that, rather than inspecting every item of a specific type—for example, portable CO2 bottles—onboard a ship or submarine, INSURV will only inspect a sample of those items. From the sample, INSURV would then characterize the ship's or submarine's entire complement of that item.

A good sampling methodology must: (1) ensure that the sample is representative of the population, and (2) the sample size is sufficiently large so that the “margin of error” of the results is small. Furthermore, a successful sampling methodology must be:

- **Defensible:** The sampling methodology must be analytically/statistically rigorous and it needs to pass the “common sense” test (so the Fleet will accept it).
- **Implementable:** It must be easy to apply by inspection teams during the inspection process. That is, the purpose of sampling is to make the inspection teams more efficient, not to add a burden to the inspection process.
- **Justifiable:** It must result in a sufficient improvement in efficiency that more than makes up for the extra complexity it will introduce.
- **Flexible:** Inspectors must have the flexibility to adjust as shipboard conditions require. In particular, the methodology cannot be so rigid that it inhibits inspectors from pursuing things they observe aboard the platform.
- **Transparent:** Inspectees will need to know that INSURV will not necessarily inspect all items, though they need to be prepared for

all, and they also will need to be prepared for, and be willing to accept, results based on sampling.

This report outlines a sampling methodology that meets the requirements of the first two bullets: it is statistically rigorous and therefore quantitatively defensible, and it is implementable. It is based on well-known sampling methods, such as those described in Cochran (1977) and Lohr (1999).

The method described herein allows INSURV to specify the desired margin of error of the results on each item. It is expected that this decision will be based on the mission essentiality and/or safety criticality of each item, where items that are mission essential or safety critical will be given very small margins of error. Similarly, items that are not mission essential or safety critical will be given appropriately larger margins of error.

Based on this choice, as well as other information about each item, such as how many are aboard a given platform, the method gives a required sample size. For mission-essential or safety-critical items, particularly when there are few of each item aboard the ship or submarine, the method often results in a requirement for 100% inspection, which is what INSURV is currently doing. However, for non-mission-essential and non-safety-critical items, particularly when there are many of the items, the sample size can be considerably smaller than the total number of items.

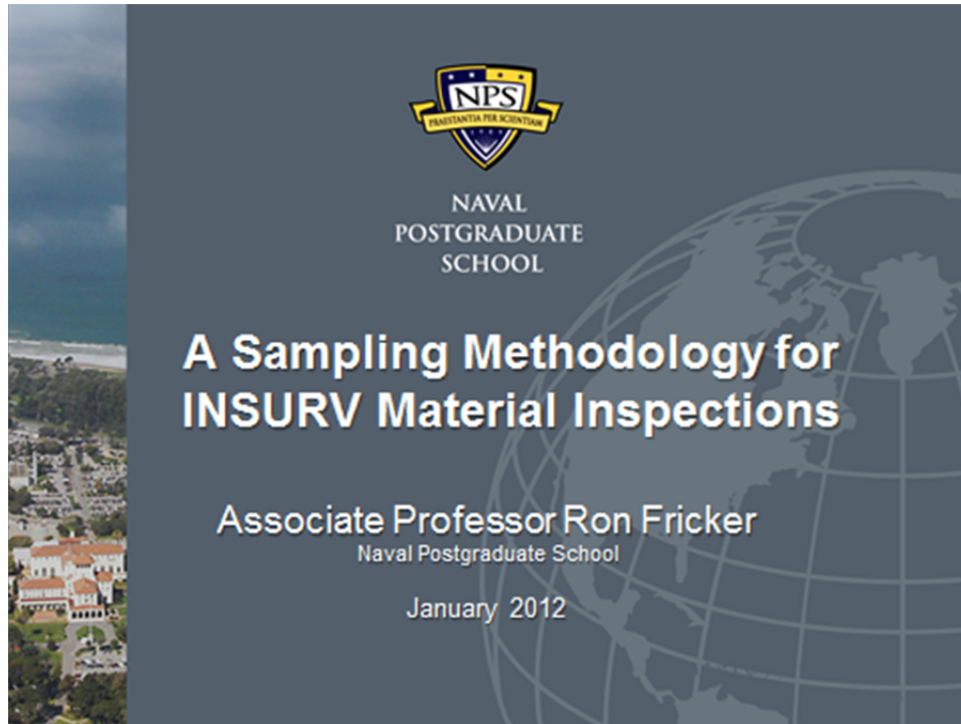
Whether the sampling method described in this report is justifiable, in the sense that the effort of implementing it is worth the benefits of sampling, is a question only INSURV can answer. If the answer is yes, then further development and training is required so that INSURV staff and inspectors can flexibly implement these methods and subsequently defend and explain the results to the Fleet.

ACKNOWLEDGEMENTS

This research was sponsored by INSURV and could not have been completed without the support of RADM Wray, President, Board of Inspection and Survey, and detailed assistance of Mr. Strait and Mr. Papis, as well as input from members of INSURV.

All errors and omissions are, of course, solely the responsibility of the author.

THIS PAGE INTENTIONALLY LEFT BLANK

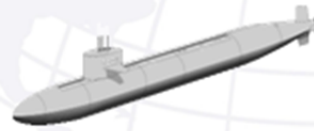


This technical report, presented in the form of an annotated briefing,¹ describes a sampling methodology that the U.S. Navy Board of Inspection and Survey (INSURV) could use for ship and submarine Material Inspections (MIs). It is largely based on a draft briefing provided to Mr. Robert Strait, Director of Operations Analysis and Chief Information Officer (CIO), Board of Inspection and Survey, in September 2011.

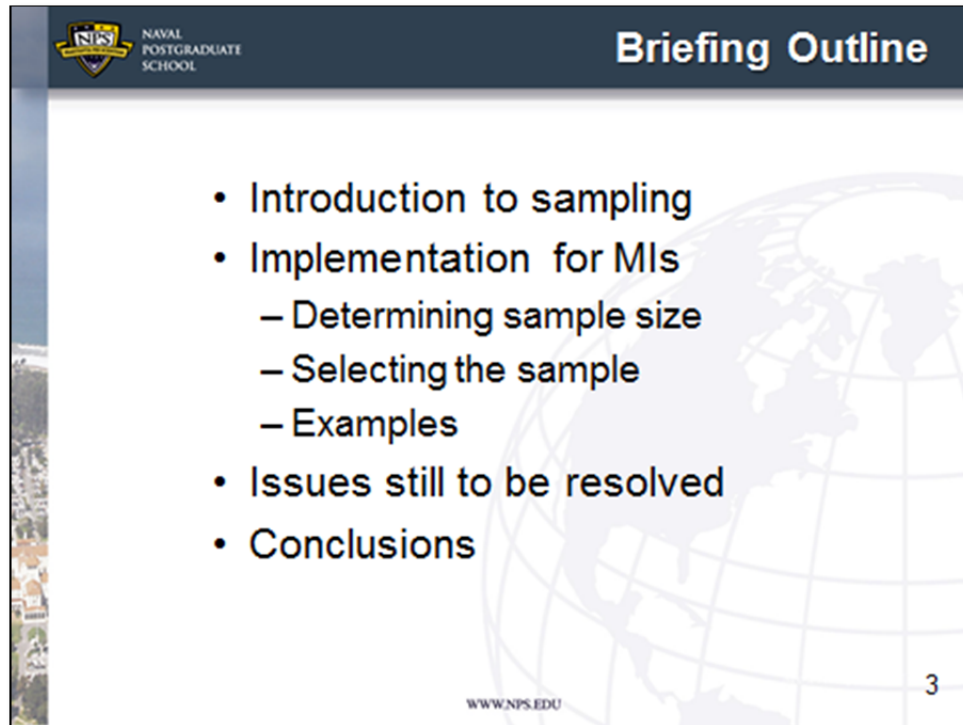
¹ As an annotated briefing, this report is intended to present in written form how the briefing would have been given verbally. The text under each slide (and sometimes continuing on to the next page) documents how each slide would have been described and presented. In so doing, the text sometimes reiterates verbiage from a slide and often amplifies and expands on the information contained in the slide.



Devise a sampling methodology
to improve
INSURV Material Inspection efficiency
without compromising
Material Inspection rigor or effectiveness

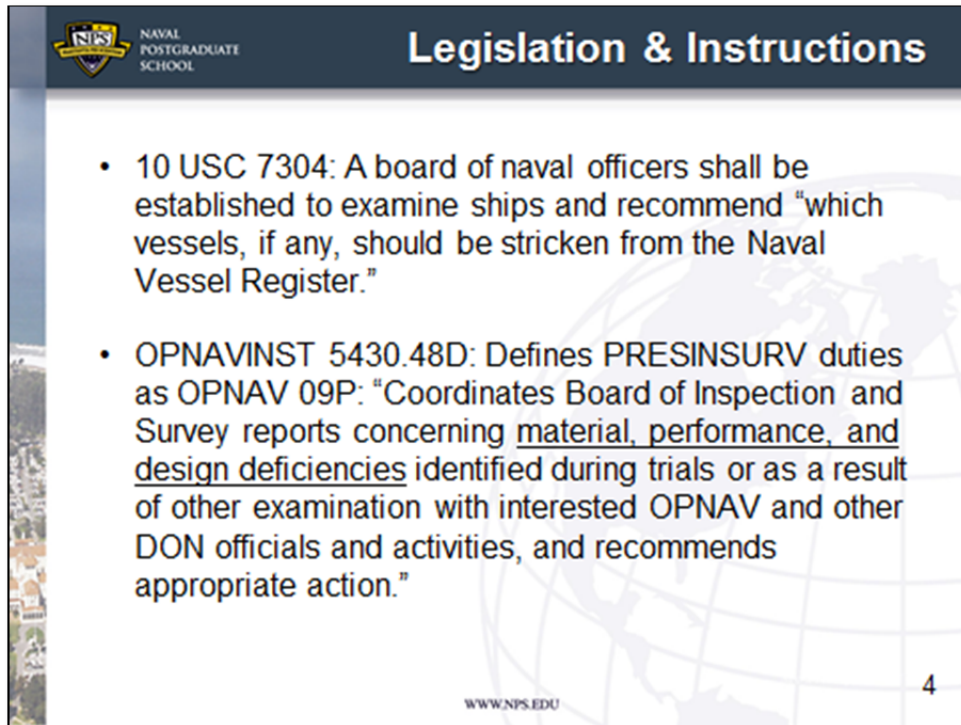



The purpose of this briefing is to present a rigorous sampling methodology that INSURV can employ to improve how the Board conducts MIs. The goal is to improve INSURV's efficiency in conducting MIs without compromising either Board effectiveness or inspection rigor.



This briefing introduces the idea of sampling, including discussing the arguments in favor of using sampling during MIs. It then describes a sampling methodology for MIs that allows INSURV to correctly and defensibly determine how many items to inspect (the sample size) as well as which items to inspect. As we will discuss, the methodology is designed to ensure samples are representative of the population and sample sizes are set so that the sampling “margin of error” is appropriately small.

That said, this report only presents an overarching sampling methodology. It does not discuss or explain how to implement the sampling methodology within INSURV’s MIs. Such implementation will require decisions about which platforms, and then which systems within platforms, sampling should be applied to. Perhaps more importantly, successful incorporation of sampling into MIs will require a well-considered implementation plan.



 NAVAL POSTGRADUATE SCHOOL

Legislation & Instructions

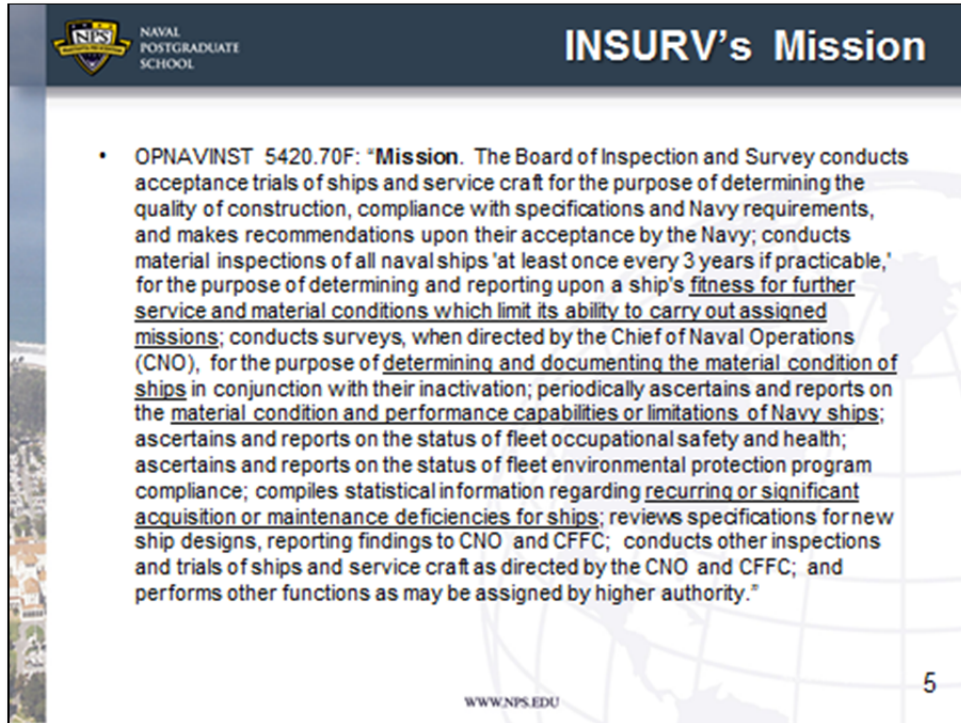
- 10 USC 7304: A board of naval officers shall be established to examine ships and recommend “which vessels, if any, should be stricken from the Naval Vessel Register.”
- OPNAVINST 5430.48D: Defines PRESINSURV duties as OPNAV 09P: “Coordinates Board of Inspection and Survey reports concerning material, performance, and design deficiencies identified during trials or as a result of other examination with interested OPNAV and other DON officials and activities, and recommends appropriate action.”

WWW.NPS.EDU 4

Before we get into the details of sampling, it is worthwhile to quickly review the relevant legislation that established INSURV and the instructions that govern it.

Originally established in 1868, in 1882 Congress enacted legislation that placed INSURV under statutory authority. Title 10 of the U.S. Code now governs the Board, though it provides little in the way of detail about how the Board should operate.

OPNAV Instruction 5430.48D, dated 29 March 1993, defines the duties of the President, Board of Inspection and Survey (PRESINSURV) as OPNAV 09P where, in particular, it says that PRESINSURV coordinates INSURV reports “concerning material, performance, and design deficiencies” and “recommends appropriate action.” Key here are the underlined words on the slide: “material, performance, and design deficiencies.”

A presentation slide titled "INSURV's Mission" from the Naval Postgraduate School. The slide features a dark blue header with the school's logo and name on the left, and the title in large white letters on the right. The main content is a bulleted list of the mission statement from OPNAVINST 5420.70F. Several phrases in the text are underlined to emphasize the mission's focus. The slide number "5" is in the bottom right corner, and the website "WWW.NPS.EDU" is at the bottom center.

NAVAL POSTGRADUATE SCHOOL

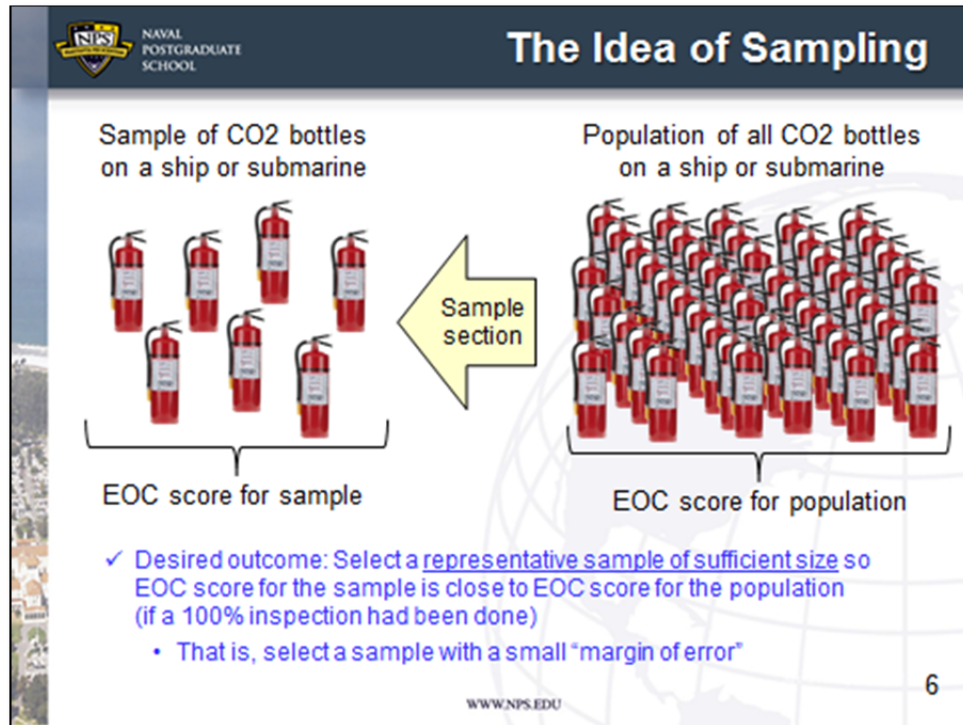
INSURV's Mission

- OPNAVINST 5420.70F: **"Mission.** The Board of Inspection and Survey conducts acceptance trials of ships and service craft for the purpose of determining the quality of construction, compliance with specifications and Navy requirements, and makes recommendations upon their acceptance by the Navy; conducts material inspections of all naval ships 'at least once every 3 years if practicable,' for the purpose of determining and reporting upon a ship's fitness for further service and material conditions which limit its ability to carry out assigned missions; conducts surveys, when directed by the Chief of Naval Operations (CNO), for the purpose of determining and documenting the material condition of ships in conjunction with their inactivation; periodically ascertains and reports on the material condition and performance capabilities or limitations of Navy ships; ascertains and reports on the status of fleet occupational safety and health; ascertains and reports on the status of fleet environmental protection program compliance; compiles statistical information regarding recurring or significant acquisition or maintenance deficiencies for ships; reviews specifications for new ship designs, reporting findings to CNO and CFFC; conducts other inspections and trials of ships and service craft as directed by the CNO and CFFC; and performs other functions as may be assigned by higher authority."

WWW.NPS.EDU

5

Finally, OPNAV Instruction 5240.70F, dated 7 August 2006, specifies INSURV's mission. The key text here is underlined, where the clear emphasis of INSURV's mission is on assessing material condition and acquisition and maintenance deficiencies. Nowhere does it say that, in the process of assessing material condition, INSURV is required to inspect every item on every platform inspected. Unquestionably, INSURV must correctly, accurately, and defensibly determine the material condition of a platform, but how INSURV accomplishes this is seemingly within the purview of PRESINSURV.



The idea of a sampling is that, rather than inspecting every item of a specific type—such as CO2 bottles onboard a ship or submarine—INSURV will only inspect a sample of those items. Inherent in this idea is that there is sufficient information in the sample from which INSURV can make an accurate determination about the material condition of the entire platform with respect to the item.

Of course, the measurement resulting from the examination of a sample (e.g., the CO2 bottle EOC score) is unlikely to exactly match the measurement that would have occurred had the whole population been inspected. Thus, the goal of a good sampling methodology is to ensure that the EOC score estimated from the sample is “close” to the EOC score that would have been determined if the entire population had been inspected.

As we will discuss in this report, there are two critical sampling criteria:

- (1) the sample must be representative of the population, and
- (2) the sample size must be sufficiently large so that the sampling “margin of error” is appropriately small.

By ensuring these two criteria are achieved, it is possible to have confidence that the result from inspecting the sample of items does, indeed, reflect the entire shipboard complement of items. We will return to these ideas, first more rigorously defining them, and then discussing how a sampling methodology can be implemented that meet the criteria.



What's the Margin of Error?

- **Margin of error** quantifies the uncertainty inherent in using a sample to “make inference” about a population
 - Technically, the margin of error is the half-width of a 95% confidence interval around the statistic of interest
- For MIs, add and subtract the margin of error from the sample EOC to get an interval within which the population EOC lies with high confidence

The “margin of error” quantifies the uncertainty inherent in sampling. When used in the context of polling results, the margin of error is typically taken to mean that the result for the population is highly likely to be within the observed result of the poll, plus or minus the margin of error. While not technically correct, it does capture the idea that the margin of error is a measure of the uncertainty inherent in the sample.

Technically, the margin of error is the half-width of a 95% confidence interval around the statistic of interest. So, adding and subtracting the margin of error from the EOC score of the sample, say for the CO₂ bottles from the previous slide, gives an interval within which the population EOC score lies with high confidence.

Perhaps most relevant to this discussion, note that the margin of error quantifies the uncertainty in the result. The larger the margin of error, the more likely the result observed from the sample can deviate significantly from what would have been observed if the entire population of items had been inspected. The good news is that, via the sample size, INSURV can control the margin of error of its results.



Margin of Error Examples

- **National polls example:** A Rasmussen Reports poll says 42% of 1,000 likely voters rate the President as a good or excellent leader*
 - Poll has a 3% margin of error
 - So we can be highly confident that between 39 and 45% of likely voters in the US rate the President as a good or excellent leader
- **MI example:** After inspecting 20 of 100 CO₂ bottles on USS Antietam, estimate EOC of the population of 100 bottles as 0.85 ± 0.03
 - Highly confident ship is SAT for CO₂ bottles

* http://www.rasmussenreports.com/public_content/politics/obama_administration/september_2011/42_give_obama_positive_marks_on_leadership_40_give_poor_rating accessed 3 September 2011.

WWW.NPS.EDU

8

Consider two examples to illustrate the margin of error idea. First, consider a Rasmussen Reports survey of President Obama. In it, 1,000 likely voters were asked to rate the President, and 42% of them said he is a good or excellent leader. The margin of error for this poll is 3%. Thus, a 95% confidence interval for the percentage of voters in the entire population who think the President is a good or excellent leader is from 39% to 45%. That is, we can be highly confident that between 39% and 45% of likely U.S. voters would rate the President a good or excellent leader.

For a (hypothetical) INSURV example, imagine that 50 of 100 CO₂ bottles were inspected during an MI of the USS Antietam. Further assume that the 50 CO₂ bottles are representative of the entire population, that the EOC score for the 50 bottles is 0.85, and that the margin of error is 0.03. Then a 95% confidence interval for the EOC score for Antietam's CO₂ bottles is from 0.82 to 0.88. That is, we can be highly confident that the Antietam should receive a SAT score for CO₂ bottles.

NAVAL POSTGRADUATE SCHOOL


Arguments in Favor of Sampling

- **Frees up resources**
 - Saved man-hours can also be used in other inspection areas, perhaps allowing for a deeper and/or broader assessment of a platform's material condition
 - If sampling eliminates unnecessary man-hours, then INSURV resources can be freed up for other purposes (e.g., MTIAs)
- **Informal sampling already occurs**
 - Inspectors already have/take the discretion to choose which items they inspect and to what level of detail
 - This effort simply formalizes (some) current practices and, in so doing, makes them quantitatively defensible
- **100% inspection not always necessary**
 - INSURV's mission is to assess a ship or submarine's material condition, not to document every existing deficiency
 - Overall material condition can often be accurately assessed without inspecting every single item

WWW.NPS.EDU 9

There are a number of arguments in favor of using sampling for MIs. These include:

- On a given platform, if it saves man hours that can be more effectively used in other inspection areas, it allows for a deeper and/or broader assessment of that platform's material condition.
- Across the organization, if it eliminates unnecessary man-hours, then that frees up INSURV resources that can be used in other ways (e.g., MTIAs).
- Informal sampling currently occurs, where inspectors already have/take the discretion to choose which items they inspect and to what level of detail.
- It's generally the preparation for the inspection, not the inspection itself, which results in improved material condition. Regardless, the Board's function is not to improve the material readiness of the Fleet, but rather to document the material condition.
- It's often not necessary to inspect every item to make an adequate determination of the material condition/readiness.



NAVAL
POSTGRADUATE
SCHOOL

Sampling Approaches

- Two ways to apply sampling to MIs:
 - “Within-platform” sampling: For items that occur in large quantities on particular platforms, only a subset of them may need to be inspected to determine the platform’s material condition
 - “Between-platform” sampling: For other items, it may be possible to inspect them only on some platforms and not on others
- ✓ Within-platform sampling is focus here


WWW.NPS.EDU

10

There are two ways to think about how INSURV might apply sampling. The first we’ll call “within platform” sampling, which means that the sampling will be applied to items from one ship or submarine during its MI. This is the type of sampling we’ve been discussing thus far with the CO2 bottle example.

However, “between platform” sampling may also be of interest. Here the idea is that certain items are only inspected on *some* ships or submarines, not all of them. Obviously, this type of sampling is not relevant to MIs, but it could be useful if INSURV was tasked with assessing Fleet readiness for some item or program that is not part of an MI. Under these conditions, it may not be necessary to inspect every item on every platform, but rather only a sample of platforms.

In any case, since this report is concerned with how sampling might apply to MIs, we will only concern ourselves with within platform sampling.



NAVAL
POSTGRADUATE
SCHOOL

Sampling Principles

- **Defensible:** Sampling methodology needs to be analytically/statistically rigorous and it needs to pass the "common sense" test (so the fleet will accept it)
- **Implementable:** It must be easy to apply by inspection teams during the inspection process
- **Flexible:** Inspectors must have the flexibility to adjust as shipboard conditions require
- **Justifiable:** It must be explainable to the Fleet and result in efficiency gains that more than make up for the extra complexity it will introduce
- **Transparent:** Inspectees need to know that INSURV will not necessarily inspect all items, though they need to be prepared for all

WWW.NPS.EDU

11

Before continuing, it is important to lay out the principles that a sampling scheme should meet if it is to be applied by INSURV to MIs. Such a sampling scheme must be:

- **Defensible:** The sampling methodology must be analytically/statistically rigorous *and* it needs to pass the "common sense" test (so the Fleet will accept it).
- **Implementable:** It must be easy to apply by inspection teams during the inspection process. That is, one purpose of sampling is to make the inspection teams more efficient, not to add a burden to the inspection process.
- **Flexible:** Inspectors must have the flexibility to adjust as conditions on the ground require. In particular, the methodology cannot be so rigid that it inhibits inspectors from pursuing things they observe during an MI.
- **Justifiable:** It must result in a sufficient improvement in efficiency that more than makes up for the extra complexity it will introduce. Furthermore, from an INSURV organizational standpoint, sampling

is only a benefit if it eliminates resources that are required for material inspections.

- **Transparent:** Inspectees need to know that INSURV will not necessarily inspect all items, though they need to be prepared for all. They also will need to be prepared for, and be willing to accept, results based on sampling.



Critical Considerations

- With sampling, it is critical to ensure:
 1. The sample size is appropriately large, and
 2. The sample is representative of the population
- We'll discuss each of these in turn...

As previously mentioned, there are two critical criteria for valid sampling:

- First, the sample must be representative of the population. That is, it cannot be focused on any particular subset, say by location, or division. Representativeness is usually achieved by randomly sampling from among all the items aboard a platform (i.e., the “population”) according to specific guidelines that we will discuss.
- Second, the sample must also be of sufficiently large size to ensure an appropriately small “margin of error.” The margin of error itself should be a function of the criticality of the item being inspected. Thus, items with higher criticality should also have larger sample sizes, while items with lower criticality should have smaller sample sizes.

Much of the rest of this report will be about how to determine the necessary sample size and then how to select a representative sample.



About Sample Size

- Correct sample size is dependent on:
 - Size of the population,
 - Inherent variation in the population, and
 - Desired margin of error
- Only one of these quantities is within INSURV's control: margin of error
 - Population size (N) is what exists shipboard
 - Population variation (measured in terms of EOC standard deviation, s) is a function of how consistently the items are maintained

The correct sample size for a particular inspection evolution is dependent on three quantities. First, it is dependent on the total number of items that could be inspected, which we call the size of the population. Second, it is dependent on the measurement variation in the population, meaning how much the items themselves vary in terms of the measurements being taken or observed. And, third, the sample size depends on the desired margin of error.

Of these, only the margin of error is within INSURV's control (indeed, it should be specified by INSURV). For the other two, the population size (which is usually denoted by a capital letter N) is simply the number of items that are installed or aboard the ship or submarine being inspected. The variation in the population, measured in terms of the standard deviation—usually denoted by σ —of the EOC score for that item, is a function of how that particular item is operated, maintained, and probably a host of other factors.

The key point is that N and σ are simply a function of what exist onboard the platform being inspected. The population size N , at least for most items, should be known precisely before the MI, while the EOC standard deviation will have to be estimated, probably from historical data.

In particular, the population standard deviation σ is usually estimated by the sample standard deviation, denoted by s , and it is calculated in the usual way. That is, if x_1, x_2, \dots, x_n are the observed EOC scores on a sample of n items, then

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

where \bar{x} is the sample average.



Population Variation (s)

- For items with higher variation, larger sample size required to achieve same margin of error (all else being equal)
- To simplify implementation, sample sizes calculated for four std. deviation levels
 - To use, estimate EOC standard deviation s , then use category with next higher s value
 - Or calculate per slide 19

Population EOC Variation	Standard Deviation
Low	$s \approx 0.1$
Moderate	$s \approx 0.2$
High	$s \approx 0.3$
Very High	$s \approx 0.4$

An important point to note is that, for a fixed population size N , populations that are more variable (i.e., have greater EOC standard deviations) will require larger sample sizes to achieve a desired margin of error (compared to an equivalent population with a smaller EOC standard deviation).

Now, while s is continuous and, at least theoretically can take on any nonnegative value, for the purposes of the calculations in this briefing, we will discretize s into four levels and base the sample size calculations on these four levels. As shown in the slide, we call these levels “low,” “moderate,” “high,” and “very high” population variations. For the purposes of this report, they are defined as EOC standard deviations of 0.1, 0.2, 0.3, and 0.4, respectively.

These EOC standard deviation levels were derived from a brief analysis of some MI data for the CG-47 class.

Alternatively, INSURV can use the precise values of s for each item. This will require using the equation on Slide 19 to determine the required sample size, rather than the tabulated values we will discuss shortly.

Should INSURV decide to implement a sampling scheme, a more rigorous analysis of these levels should be undertaken to ensure they are the most

relevant. Then, during implementation, they should be used conservatively, meaning that sample size determinations should be based on rounding up to the largest reasonable value of s .



Margin of Error (m)

- **Smaller margins of error require larger samples**
 - Zero margin of error requires 100% inspection
 - This is where INSURV currently operates
- **Item criticality should drive margin of error**
 - Mission critical/safety → smaller margin of error
 - Non-mission critical → larger margin of error
- **To simplify, calcs use four margin of error categories**
 - If another margin of error desired, calculate per slide 19

Item Criticality	Margin of Error
Very low	0.01 or 1%
Low	0.03 or 3%
Moderate	0.05 or 5%
High	0.1 or 10%

15


WWW.NPS.EDU

As we discussed, the sample size is a function of the margin of error, which INSURV would set. All other things being equal, a smaller margin of error will result in a larger sample size. At the extreme, a zero margin of error results in 100% inspection—i.e., all the items in the population must be inspected. This is where INSURV currently operates.

However, a zero margin of error may not be necessary for all inspection items and, in fact, the margin of error should be driven by the criticality of the item being inspected. That is, for mission-critical items, where it is vital that INSURV measures the population with a high degree of accuracy, there should be a small margin of error (resulting in a large sample size, which could be the entire population). On the other hand, for non-mission-critical items, a larger margin of error may be acceptable, which could result in not all the items being inspected.

To simplify the sample size calculations, we use four margins of error levels, which we characterize as “very low,” “low,” “moderate,” and “high.” These correspond to margins of error of 1%, 3%, 5%, and 10%, respectively.

As with the EOC standard deviation, INSURV does not have to use these particular margin of error levels. Doing so facilitates the use of the tables that we will discuss shortly (on Slide 18), but the equation on Slide 19 can be used to determine the required sample size for any desired margin of error.



NAVAL
POSTGRADUATE
SCHOOL

Some Illustrative Examples

- Population variation*
 - Low: "Crew berthing" median $s=0.08$; "AN/SLQ-32" median $s=0.12$; "WTD" median $s=0.13$; "Fire pump" median $s=0.14$
 - Moderate: "Fan room" median $s=0.19$; "VLS" median $s=0.20$; "Radio" median $s=0.20$; "CHT" median $s=0.24$
 - High: "Ventilation" median $s=0.29$
 - Very high: Perhaps none?
- Margins of error
 - Very low: AN/SPS-49V8, magazine sprinklers
 - Low: watertight doors, power panels
 - Moderate: portable fire extinguishers, telephones
 - High: rack lights, laundry equipment

* Calculated from MIDATA Warehouse using CG-47 MI data from years 2009-2011 by searching on key word, calculating the EOC standard deviation for each hull, and then reporting the median EOC standard deviation. However, note that this is *not* the right data to use since it is only comprised of deficiencies. The correct calculation should use information on the entire population which includes both deficient and non-deficient items. Thus, these examples may be underestimating the total population variation.

WWW.NPS.EDU

16

This slide shows some examples of population variation and possible margins of error. The population variation (i.e., the EOC standard deviations) were calculated from the MIDATA Warehouse using CG-47 MI data from years 2009-2011 by searching on key word, calculating the EOC standard deviation for each hull, and then reporting the median EOC standard deviation.

However, it is important to note that this is *not* the right data to use since it is only comprised of deficiencies. The correct calculation should use information on the entire population, which includes both deficient and nondeficient items. Thus, these examples are likely underestimating the total population variation. Nonetheless, they are included here for illustrative purposes.

The margins of error are also illustrative, but they show how mission-critical items such as mission-essential and safety equipment should have low margins of error, while non-mission-essential equipment, such as laundry equipment, should have larger margins of error.



Determining the Sample Size

1. Get population size (N) and variability (s)
 - From class knowledge and historical data
 - If unsure, err conservatively: overestimate N and/or s
2. Decide on margin of error category
 - How critical is the item being assessed?
3. Look up required sample size in the tables on slide 18
 - Interpolate or calculate directly (see slide 19) as desired/necessary

17

WWW.NPS.EDU

So, with that as background, how to calculate the sample size? There are three steps in the process. First, determine the population size N and the variability s of the item. These quantities follow from class knowledge and historical data. Second, decide on the margin of error. This is simply INSURV's choice based on the criticality of the inspection item. Finally, look up the required sample size in the table on the next slide.



Required Sample Sizes Tables

Low population variability ($s=0.1$)		Population Size (N)														
Margin of Error (m)		1	2	3	4	5	10	15	20	25	50	100	150	200	250	500
Very low: 0.01		1	2	3	4	5	10	15	20	24	45	80	110	134	154	223
Low: 0.03		1	2	3	4	5	9	12	14	16	24	31	35	37	38	41
Moderate: 0.05		1	2	3	4	4	7	8	9	10	13	14	15	15	16	16
High: 0.1		1	2	2	2	3	3	4	4	4	4	4	4	4	4	4

Moderate population variability ($s=0.2$)		Population Size (N)														
Margin of Error (m)		1	2	3	4	5	10	15	20	25	50	100	150	200	250	500
Very low: 0.01		1	2	3	4	5	10	15	20	25	49	95	138	178	217	381
Low: 0.03		1	2	3	4	5	10	14	18	22	40	64	82	95	104	132
Moderate: 0.05		1	2	3	4	5	9	13	16	18	29	40	45	49	51	57
High: 0.1		1	2	3	4	4	7	8	9	10	13	14	15	15	16	16

High population variability ($s=0.3$)		Population Size (N)														
Margin of Error (m)		1	2	3	4	5	10	15	20	25	50	100	150	200	250	500
Very low: 0.01		1	2	3	4	5	10	15	20	25	50	98	144	190	234	440
Low: 0.03		1	2	3	4	5	10	15	20	24	45	80	110	134	154	223
Moderate: 0.05		1	2	3	4	5	10	14	18	22	38	60	74	84	92	112
High: 0.1		1	2	3	4	5	8	11	13	15	21	27	30	31	32	34

Very high population variability ($s=0.4$)		Population Size (N)														
Margin of Error (m)		1	2	3	4	5	10	15	20	25	50	100	150	200	250	500
Very low: 0.01		1	2	3	4	5	10	15	20	25	50	99	147	194	241	464
Low: 0.03		1	2	3	4	5	10	15	20	25	47	88	124	157	185	294
Moderate: 0.05		1	2	3	4	5	10	15	19	23	42	72	95	113	127	170
High: 0.1		1	2	3	4	5	9	13	16	18	29	40	45	49	51	57

WWW.NPS.EDU

18

This slide is the table of required sample sizes, one for each level of population variability s . Given s , each table is a function of the margin of error m and the population size N . Simply enter the appropriate table and read off the sample size for the population size N and the desired margin of error m .

For example, consider an item with an EOC variation of $s=0.4$ ("very high"), a population size of $N=100$, and a desired margin of error of $m=0.1$ (10% or "high"). For this item the required sample size is 40. That is, under these conditions, only 40% of the items need to be inspected. In contrast, for an item with an EOC variation of $s=0.1$ ("low"), a population size of $N=10$, and a margin of error of $m=0.01$ (1% or "very low"), the required sample size is 10. That is, under these conditions, a 100% inspection is required.

For population sizes in between those listed in the tables, nonograms are provided at the end of this briefing (Slides 31-34) that give the results for any sample size up to $N=500$.



Sample Size Calculation

- Define

- N is the population size for a particular inspection item
- s is the estimated standard deviation of the EOC in the population
- m is the desired margin of error
- n is the required sample size

- Then

$$n = \left\lceil \frac{s^2 / (m/2)^2}{1 + s^2 / N (m/2)^2} \right\rceil$$

where the notation $\lceil \rceil$ means to round the quantity inside up to the next integer value

This slide shows the formula behind the sample size calculations that give the results in the tables on Slide 18 (and the nomograms on Slides 31-34). Note how the sample size is related to the various parameters s , N , and m :

- As the standard deviation in the population s gets larger (all else staying constant), a larger sample size n is required to achieve the same margin of error.
- Similarly, as the population size N increases (all else staying constant), a larger sample size n is required to achieve the same margin of error.
- To achieve a smaller margin of error m (again, holding all else constant), a larger sample size n is required.



- Knowing how many items to sample is only half of the solution – also need to know which items to inspect
- Critical consideration is sample must be representative of shipboard population
 - Intuitively, idea is to ensure that
 - Every item has a chance of being inspected, and
 - Inspected items are selected throughout the ship
- Sample selection must be based on systematic approach, not inspector judgment
 - Only way to defend against allegations of favoritism or bias

As we discussed earlier, there are two important considerations in implementing a rigorous sampling scheme. One is the determination of the sample size. The second is the methodology for selecting a representative sample. Intuitively, a representative sample must allow each item to be in the sample. That is, it cannot arbitrarily eliminate specific items from being sampled. In addition, the methodology should ensure that the sample comes from throughout the ship or platform.

A critical consideration is that the sampling methodology must eliminate human judgment from the sample selection process. There is an important reason for this, namely that if the sample selection is left up to human judgment, then it is possible for conscious or subconscious biases to creep into the sample. For example, an inspector may tend to focus on items that are likely to be defective, thereby lowering the estimated EOC. Or perhaps another inspector decides to only sample easily available items, which also just happened to have better maintenance for the same reason, thereby overestimating the EOC.

These statistical issues aside, it's also the only way to defend against allegations that the inspector chose a biased sample, biased either for or against

the platform. That is, a formal methodology for selecting the sample that is outside the inspector's control also gives the inspectors a defense against allegations of either favoritism or antagonism.



Systematic Sampling

- Selection of items to inspect based on technique called **systematic sampling**
- If $n \leq N/2$, calculate $k = \lfloor N/n \rfloor$, choose a random item to start with and then inspect every k^{th} item after it
 - Notation: round N/n down to next integer
 - Example: Base security sometimes inspects every k^{th} vehicle coming through the gate
- Else, if $n > N/2$, calculate $k = \lceil N/(N-n) \rceil$, choose a random item to start, and skip every k^{th} item after it
 - Notation: round $N/(N-n)$ up to next integer

A sampling strategy that meets the criteria on the previous slide and is also easy to implement is called *systematic sampling*. A simple example will make the idea clear. Imagine a population of $N=100$ items with a required sample size of $n=10$. A systematic sample results if the inspector chooses a random item to start with and then inspects every tenth item after it.

The calculation gets a little more complicated than that if the population size divided by the sample size (N/n) is not an integer, or if n is more than 50% of N . In the former case, the solution, as shown in the slide above, is to round N/n down to the next lowest integer. For example, if $n=10.2$, simply round it down to 10.

In the latter case, the calculation is shown in the last bullet of the slide above, where the idea is to determine which items to skip and not inspect. As a simple example, imagine the case where $N=9$ and $n=6$. Then $N/(N-n)=3$ and every third item would be skipped and not inspected. For example, if the items are numbered 1,2,...,9 and the inspector picked item 5 to start, then he or she would inspect items 5, 6, 8, 9, 2, and 3, skipping over items 7, 1, and 4.



Example #1: Enlisted Berthing

- Goal: Determine the number of racks to inspect in one berthing compartment in support of Event ID HAB0010
- Scenario:
 - Compartment has $N=50$ racks
 - Data from similar ships gives EOC standard deviation $s=0.1$ ("low variability")
 - Desired margin of error "moderate" ($m=0.05$)
- ✓ Then, required sample size is $n=13$ and sample selection is to inspect every third rack

Note: All data in this example is hypothetical and purely for illustrative purposes

WWW.NPS.EDU

22

Let's look at an example that brings it all together. In this case, it's the inspection of racks in one berthing compartment in support of Event ID HAB0010. The scenario is that the compartment has $N=50$ racks, for which an analysis of data from previous ships of the same class gives an EOC standard deviation of $s=0.1$, and INSURV determines that the margin of error for this inspection item should be "moderate," so $m=0.05$.

Under these conditions, the table on Slide 18 gives a sample size of $n=13$ racks. Thus, we have that $N/n=50/13=3.8$. Rounding this down gives $k=3$, so the inspector would choose a random rack to start with and then inspect every third one after that.



Example #2: Watertight Doors

- Goal: Determine the number of watertight doors to inspect throughout ship in support of Event ID DC0350
- Scenario:
 - $N=200$ watertight doors, scuttles and hatches
 - Data from similar ships gives EOC standard deviation $s=0.2$ (“moderate variability”)
 - Desired margin of error “low” ($m=0.03$)
- ✓ Then, required sample size is $n=95$ and sample selected by inspecting every other watertight door, hatch, or scuttle

Note: All data in this example is hypothetical and purely for illustrative purposes

WWW.NPS.EDU

23

Here’s another example, where the Event is ID DC0350, to inspect the watertight doors throughout the platform. In this scenario, there are $N=200$ watertight doors, for which an analysis of data from previous ships of the same class gives an EOC standard deviation of $s=0.2$, and INSURV determines that the margin of error for this inspection item should be “low,” so $m=0.03$.

Under these conditions, the table on Slide 18 gives a sample size of $n=95$ doors. Thus, we have that $N/n=200/95=2.1$. Rounding this down gives $k=2$, so the inspector would choose a random watertight door to start with and then inspect every other one after that.



Example #3: CWIS


- Goal: Determine the number of CIWS to inspect in support of Event ID WP0130
- Scenario:
 - Ship class has $N=2$ CIWS
 - Data from similar ships gives EOC standard deviation $s=0.3$ (“high variability”)
 - Desired margin of error “very low” ($m=0.01$)
- ✓ Then, required sample size is $n=2$ and “sample selection” is a 100% inspection

Note: All data in this example is hypothetical and purely for illustrative purposes

WWW.NPS.EDU

24

As a third and final example, consider the Event is ID WP0130, to inspect the CIWS. In this scenario, there are $N=2$ CIWS, for which an analysis of data from previous ships of the same class gives an EOC standard deviation of $s=0.3$, and INSURV determines that the margin of error for this inspection item should be “very low,” so $m=0.01$. Under these conditions, the table on Slide 18 gives a sample size of $n=2$, so the inspector would do a 100% inspection.

NAVAL
POSTGRADUATE
SCHOOL

Some Important Issues
to be Resolved

1. Items need to have some ordering to be able to determine the next k^{th} item
 - E.g., how to select “every other” water tight door distributed throughout a ship?
2. Need to avoid situations where k also corresponds to some arrangement of the items
 - E.g., for inspection of racks that come in tiers of three, inspecting “every third rack” could end up with only top racks inspected

WWW.NPS.EDU


25

Now, there are some real-world complications that need to be addressed in the implementation of systematic sampling. The first is that the items have to have some type of natural ordering in order to be able to inspect every k^{th} item. For example, what does it mean to an inspector on the deckplates to inspect “every other” watertight door? To implement such a sampling scheme, it would probably be necessary to devise some type of route around the ship or submarine over which the inspector would stop and inspect every other watertight door as he or she comes to it on the route.

The second complication is that the frequency of items to inspect (or skip) must not match up with some systematic feature of the items being inspected. In the rack inspection example, for example, racks often come in tiers of three, so inspecting “every third” rack could result in only top or bottom racks being inspected.

In this case, a solution might be that the inspector has to randomly choose one rack in each tier to inspect, though a system would need to be put in place to remove subjectivity from the choice. Alternatively, the inspector might simply systematically subsample the racks in each tier, perhaps starting with the top

rack in the first tier inspected, proceeding to the middle rack in the second tier, to the bottom rack in the third tier, and then repeating this pattern for the fourth, fifth, and sixth tiers, etc.



NAVAL
POSTGRADUATE
SCHOOL

Some Important Issues to be Resolved (continued)

3. What if the resulting interval straddles degraded or UNSAT category?
 - Could either leave it to the inspection team to determine final status or could require second sample to decrease margin of error
 - Former consistent with current practice; latter likely to be unworkable in the field
4. How much of the sampling calculations should be left to the inspectors versus being done ahead of time for them?

WWW.NPS.EDU

26

A third complication arises if the resulting interval (the EOC score of the sample plus and minus the margin of error) straddles two categories, say degraded and unsat (e.g., the interval turns out to be 77%-83%). It's not clear, then, from the results of the sample, whether the item should be given a degraded or unsat rating.

There are two possible solutions for this. The first is to let the inspection team make a subjective call as to how the item is to be rated. The second is to collect more data in order to decrease the margin of error, with the hope that the new interval will fall entirely in one rating category or the other.

The former is consistent with current practice and is probably the appropriate solution. However, it does open the inspection team up to second guessing and allegations of bias. On the other hand, the latter is likely to be unworkable in the field, as it would require the inspectors and ship's force to revisit an inspection already seemingly completed. Furthermore, it's still possible that the new interval will straddle the two rating categories (unless the second inspection results in a 100% inspection).

A fourth issue that would need to be addressed is the question of how much of the sampling calculations should be done by the inspectors during the inspection versus how much should be done for the inspectors ahead of time.



Conclusions

- Via the sampling methodology presented in this report, INSURV can explicitly trade off level of inspection effort versus inspection precision
- Returning to the sampling principles, the methodology is:
 - Statistically **rigorous** and thus quantitatively **defensible**
 - Whether INSURV feels it is **justifiable** to the Fleet and, if implemented, how **transparent** INSURV wants to be with it are yet to be determined
 - Work is still required to ensure it is appropriately **flexible** for INSURV inspectors

27

WWW.NPS.EDU

This report has presented a sampling methodology that is statistically rigorous and quantitatively defensible. What it does is allow INSURV to explicitly trade off the desired inspection precision of a system against the required inspection level of effort. The current inspection methodology—100% inspection—also implicitly makes such a trade-off, requiring maximum precision for all systems regardless of whether such a level of precision is necessary or appropriate.

Of course, just because the method is quantitatively defensible does not mean it is justifiable to the Fleet, or that the Fleet will find it acceptable. This is not something that can be addressed via statistics; it is a substantive issue that only PRESINSURV and Fleet leadership can determine. And, while the methodology is theoretically feasible, additional work is required to understand whether there are implementation issues that will need to be addressed. We briefly address some of these issues in the next three slides.



From Theory to Real World

- This report only presents the overarching methodology, not how to implement
 - Which platforms? Which systems?
 - Good implementation critical to success
 - Also “important issues to be resolved”
- Careful, measured, and well-considered roll-out plan is necessary to ensure successful implementation
 - Devil is in the details of how to actually implement and execute on the deckplates

First, it is important to emphasize that this report only presents an overarching sampling methodology. It does not discuss or explain how to implement the sampling methodology within INSURV's MIs. Such implementation will require decisions about which platforms, and then which systems within platforms, sampling should be applied to. Perhaps more importantly, successful incorporation of sampling into MIs will require a well-considered implementation plan. That is, it will be critical to get all deckplate implementation and execution details exactly right before a full roll-out to the Fleet.



Possible Next Steps

- Should INSURV decide to continue, consider conducting a beta test on one or two (habitability?) systems
- Must assess how methodology works (or not) under actual conditions; e.g.,
 - Organizational responsibilities for design, implementation, and execution
 - How to execute systematic sampling shipboard?
 - How much flexibility do inspectors require?

Given all the foregoing, should PRESINSURV determine that the sampling methodology is still worth pursuing, the next logical step is a limited beta test on a small number of non-mission-critical, non-safety systems. Some of the habitability inspections seem like good potential candidates. The issue is to assess how sampling would work under actual shipboard conditions and to uncover any real-world impediments to the use of sampling. For example, as described in Slides 25 and 26 (“Some Important Issues to be Resolved”), what are all the issues with and barriers to systematic sampling when applied to shipboard items? How should such systematic sampling plans be codified for the inspectors? Will they need to be tailored for each system? How much flexibility do inspectors need when executing a sampling strategy? How much support do they need?



Possible Future Steps

- Given a successful beta test, additional work required to fully implement, including:
 - As necessary, modification of methods to address any issues and complications found during beta test
 - Drafting guidelines and/or instructions, particularly for setting sample sizes (margins of error) and for conducting sampling during inspections
 - Establishing internal operating procedures and analytical support capabilities
 - Training of inspectors and support staff
 - Communication with Fleet as desired/necessary

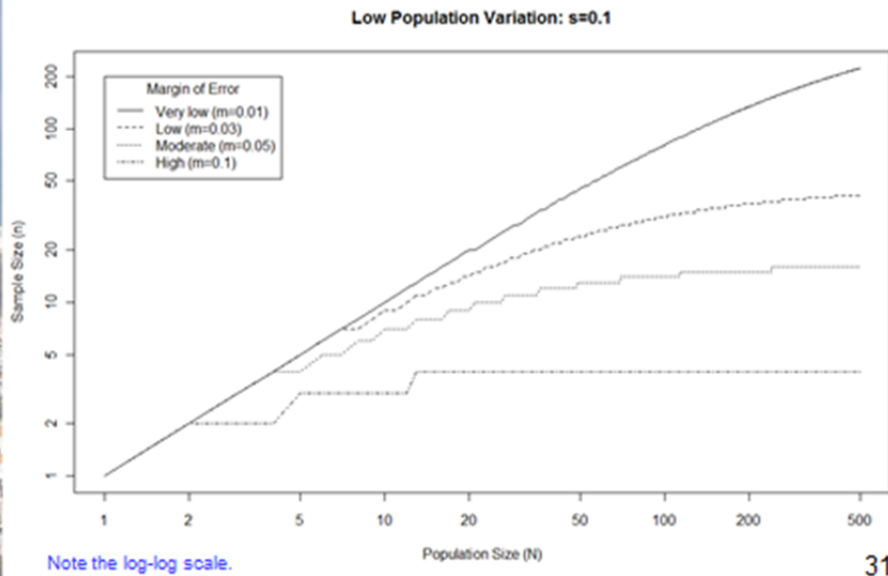
Of course, a successful beta test is only the first step in establishing sampling as a routine part of MIs. That is, not only might the methodology need to be revised to deal with any issues or complications that are found during the beta test, but a set of guidelines and/or instructions will need to be written and approved that formalize the program within INSURV. In addition to the formal guidelines, internal operating procedures will need to be established and implemented, including defining and establishing any additional analytical support capabilities that will be required. For example, INSURV will need to determine who will specify the margins of error and then who will be responsible for taking those margins of error and calculating the sample sizes and specifying sampling parameters.

Similarly, once the organizational details have been determined, resourced, and implemented, a formal training program should be conducted to educate inspectors about how to execute sampling in the field, including how to explain and justify it to ship's force (as necessary/desired). Also, INSURV will need to communicate the new policies and procedures to the Fleet and Navy

leadership. All of this will require careful planning and execution to ensure the changes are understood and positively received.



Back-up Slide: Sample Size Nomogram ($s=0.1$)

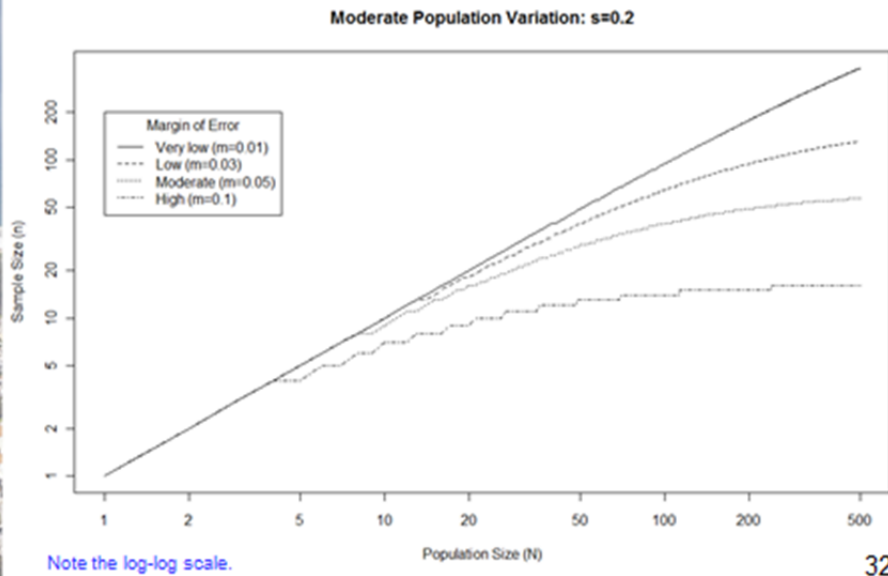


31

This is the nomogram for low population variation ($s=0.1$). It provides the sample sizes for all population sizes (up to $N=500$), not just those listed in Table 1.



Back-up Slide: Sample Size Nomogram ($s=0.2$)

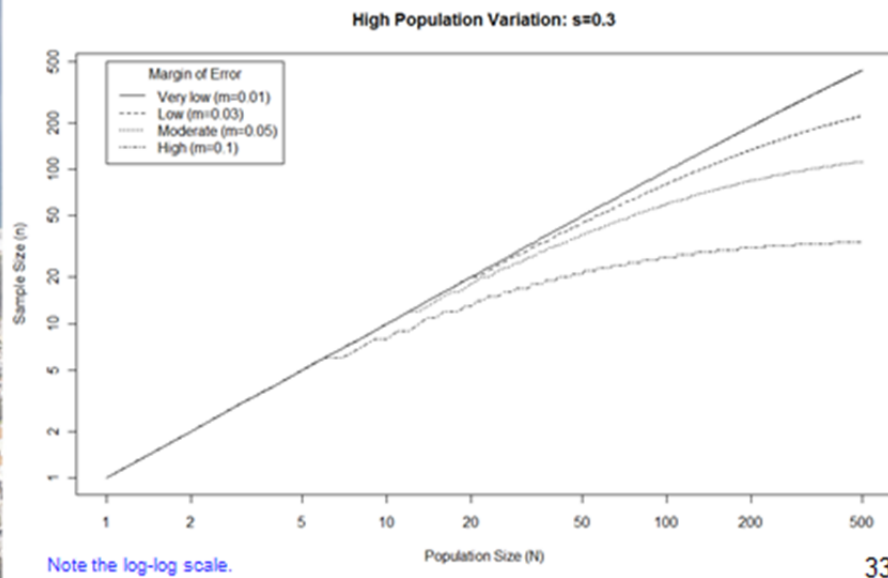


32

This is the nomogram for low population variation ($s=0.2$). It provides the sample sizes for all population sizes (up to $N=500$), not just those listed in Table 1.



Back-up Slide: Sample Size Nomogram ($s=0.3$)

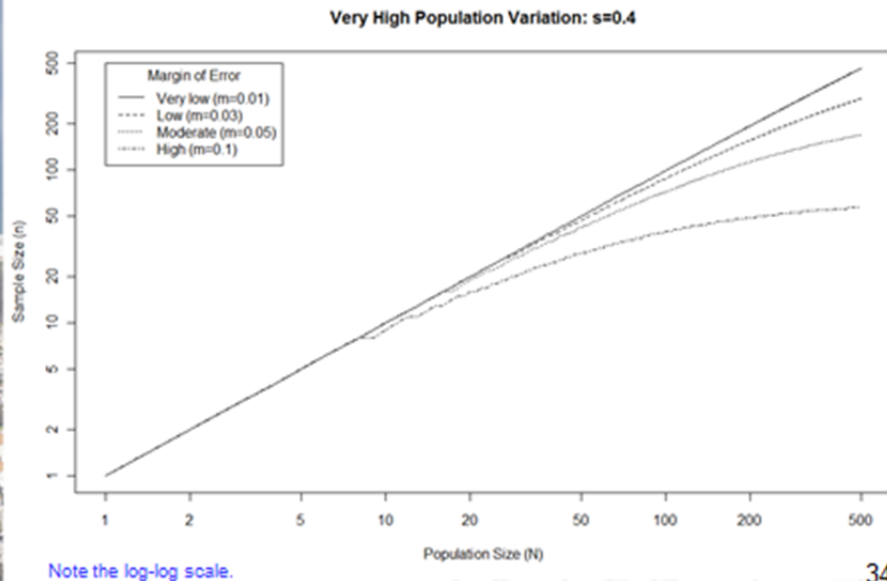


33

This is the nomogram for low population variation ($s=0.3$). It provides the sample sizes for all population sizes (up to $N=500$), not just those listed in Table 1.



Back-up Slide: Sample Size Nomogram ($s=0.4$)



34

This is the nomogram for low population variation ($s=0.4$). It provides the sample sizes for all population sizes (up to $N=500$), not just those listed in Table 1.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- Cochran, W. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.
- Lohr, S.L. (1999). *Sampling: Design and analysis* (2nd ed.). Duxbury Press.
- U.S. Navy. (1993, March 29). Office of the Chief of Naval Operations (OPNAV) organization manual. OPNAV Instruction 5430.48D.
- U.S. Navy. (2006, August 7). Mission, organizations, and functions of the Board of Inspection and Survey. OPNAV Instruction 5420.70F.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California
3. Research Sponsored Programs Office, Code 41
Naval Postgraduate School
Monterey, California
4. Richard Mastowski (Technical Editor) 1
Graduate School of Operational and Information Sciences (GSOIS)
Naval Postgraduate School
Monterey, California
5. Professor Ronald D. Fricker, Jr.2
Graduate School of Operational and Information Sciences (GSOIS)
Department of Operations Research
Naval Postgraduate School
Monterey, California
6. RADM Robert O. Wray, Jr., USN..... 1
President, Board of Inspection and Survey
2600 Tarawa Court, Suite 250
Virginia Beach, Virginia 2345
7. CAPT Thomas Snyder, USN 1
Chief of Staff
Board of Inspection and Survey
2600 Tarawa Court, Suite 250
Virginia Beach, Virginia 2345
8. CAPT Eric Tapp, USN 1
Director of Surface Inspections
Board of Inspection and Survey
2600 Tarawa Court, Suite 250
Virginia Beach, Virginia 2345
9. Mr. Robert Strait. 1
Director of Operations Analysis/CIO
Board of Inspection and Survey
2600 Tarawa Court, Suite 250
Virginia Beach, Virginia 2345